

Top 1000 Highest Grossing Hollywood Movies

A case study on Movies Collection

By Gourav Kumar

16-03-2022

Introduction

My name is Gourav Kumar and I am currently working at Cognizant as Programme Analyst. During both, my undergraduate and graduate programs, there was not any focus on Data analytics and Data Science, but I had an interest in this part of industry. So, I got myself enrolled in many courses related to this area. Those courses were the most engaging and challenging in the part of understanding. Soon I realized, that if I wanted to apply what I learned about data in the real world, and not only during exams, then I would have to take it upon myself.

So, I did a little brainstorming into what idea or problem I wanted to tackle. Pretty soon I came to the conclusion that everyone nowadays talks about Movies and how much money they have earned from it, So I decided that I should go for a data related to Hollywood movies Collection. I used to watch a lot of movies; this generates a special interest in myself.

I opted for Excel and Tableau as the primary tool of my data analysis, just because I feel so comfortable after having learned about it in the Google Data Analytics course.

Business Problem & Data

This dataset contains information about the top 1000 highest grossing Hollywood films. It is up to date as of 10th January 2022. This data has been scraped from multiple sites and has been added together for performing various data operations. The data has been taken from imdb, rotten tomatoes and many other sites. It gives the information like how much collection the movie has done Domestically as well as internationally with the Distributor name and movie duration which can be useful during the analysis.

The dataset I am going to work on is downloaded from Kaggl. It Contains 10 columns such as Title, Movie info, Distributor, Release date, Domestic sales, international sales, World sales, Genre, Movie Runtime and Licence.

Important Note

- The Domestic Collection is Separate attribute on which data is sorted.
- The World sales column is the sum of Domestic sales and international sales.

Data Cleaning

To prepare the data analysis I have cleaned the data in MS-Excel. In the Column of Release date, we had few values as NA. Since one of the analysis requires the year of release of the movie. The name of the movies contains the year of release, so for working on a complete data. I am going to put year in the missing values manually. For the proper analysis on the year wise highest grosser, I have created one column with year of release only.

Analysis

Now the data is cleaned and we can take the next step and look closer into what insights can be extracted from it. The columns "Title" and "Movie Info" have total 918 unique values. Further column analysis has been done in detail manner as follows.

Distributor

The total values in the attribute are 918, in which 34 unique entries are there and the leading distributors: the Warner Bros and Walt Disney Studios Motion Pictures with 17% occupancy each. Apart from these two others are in 64%.

Release Year

The total values are going to be 918 as well. It contains 45 unique values, which clearly means the data contains total 45 distinct years of highest sales collector movies. The Year which came maximum 42 times was 2013. In the year 2013 maximum number were released who came in the list of highest sales.

Domestic Sales (in \$)

This field contains the Domestic sales for the for the movie and it is in Dollars (\$). There is total 918 values are there and the mean for the column is \$168 million and Standard Deviation is \$104 million. The more features for the data are given in the following table.

<i>Minimum</i>	<i>80.4 million</i>
<i>25%</i>	<i>102 million</i>
<i>50%</i>	<i>132 million</i>
<i>75%</i>	<i>191 million</i>
<i>Maximum</i>	<i>937 million</i>

International Sales

This field contains the international sales for the movie and it is in Dollars (\$). There is total 918 values are there. The following table displays the features of the attributes.

<i>Mean</i>	<i>230 million</i>
<i>Minimum</i>	<i>260 million</i>
25%	88.3 million
50%	164 million
75%	289 million
<i>Maximum</i>	<i>2,090 million</i>
<i>Standard Deviation</i>	<i>217 million</i>

World Sales

This field contains the World sales for the movie and it is in Dollars (\$). There is total 918 values are there. The following table displays the features of the attributes.

<i>Mean</i>	<i>397 million</i>
<i>Minimum</i>	<i>81.6 million</i>
25%	201 million
50%	302 million
75%	475 million
<i>Maximum</i>	<i>2,850 million</i>
<i>Standard Deviation</i>	<i>308 million</i>

Genre

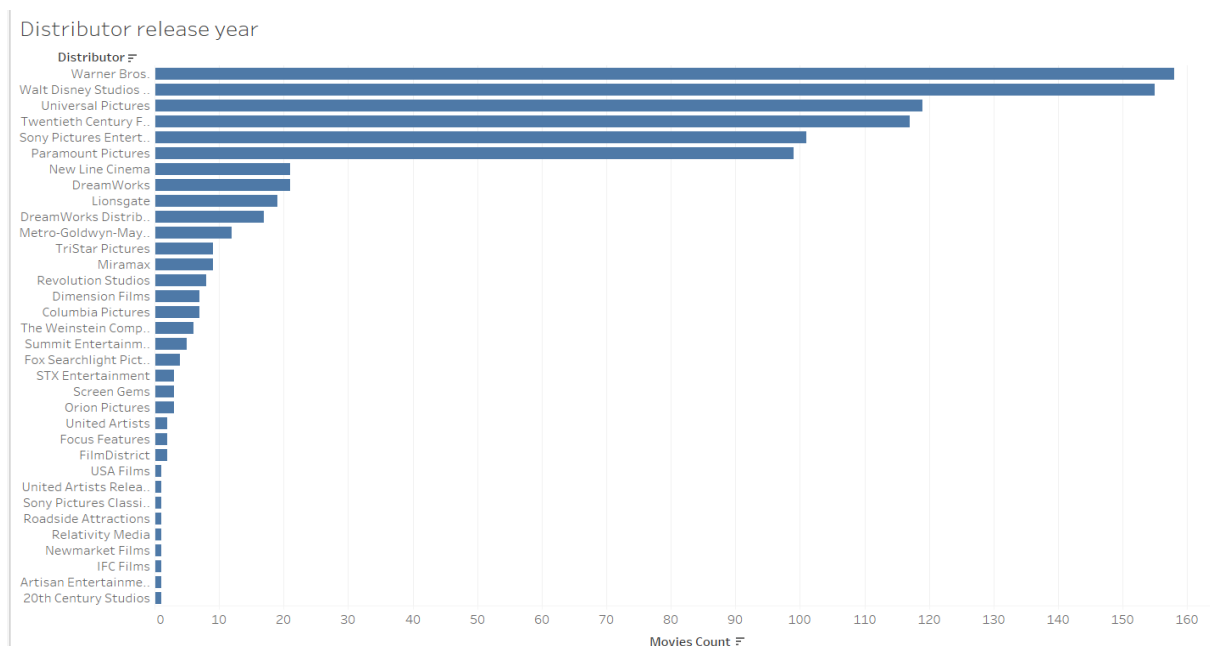
a category of artistic, musical, or literary composition characterized by a particular style, form, or content a classic of the gothic novel genre. There is total 918 values combining total 309 unique values. The most common genre with 6% occurrence is ['Action', 'Adventure', 'Sci-Fi'] and with 3% of occurrence ['Comedy'] comes at second place.

Movie Runtime

Runtime is the time between the starting of the movie upto the end of the credits scene. It does not include any trailers and break; it depends upon the theatres to put trailers and interval. There is total 918 values combining total 101 unique values. The most common Movie runtime is 1 hr 55 min. with 3% occurrence and 1 hr 56 min. with 3% occurrence. The rest 94% combined other timings.

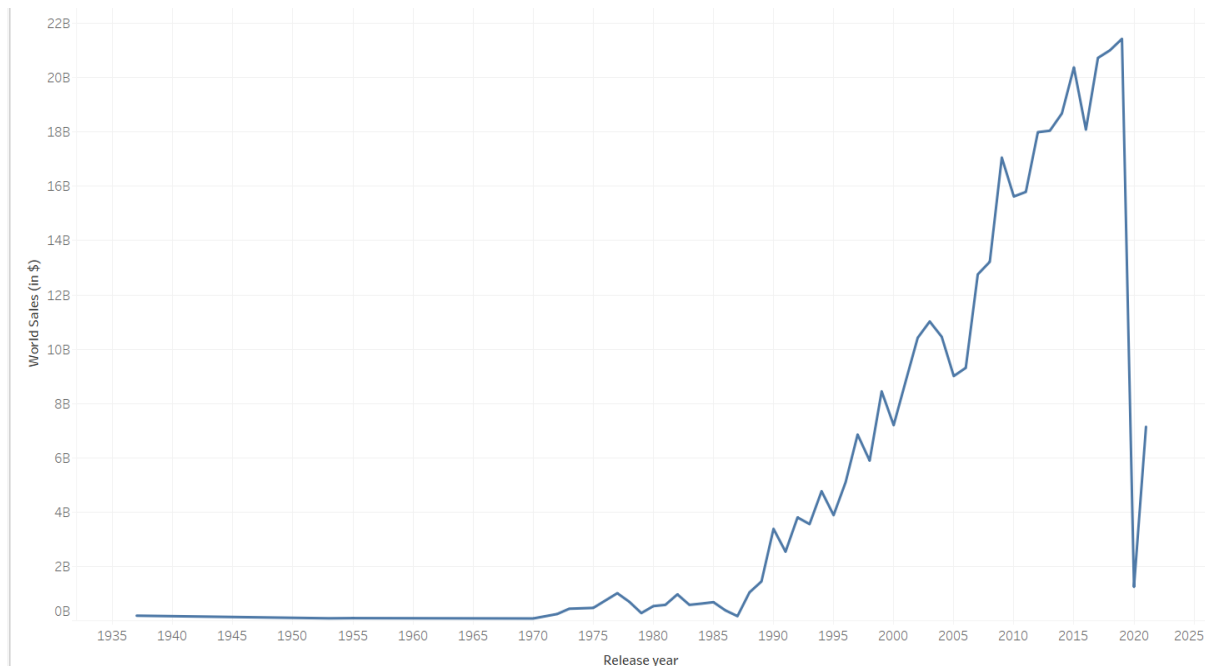
Tableau Analysis

The movies that are in this dataset combines of the highest earning 918 movies. There is total 34 Distributors. There are certain number of movies that has been created during the time period of 1937 to 2022. Following is the graph that describes the number of movies created by each distributor.



Warner Bros total 158 movies were in the highest grossing list and on the second position we have Walt Disney Studios motion picture with 155 movies. Out of 34, the six Distributors had only one movie in the list. Although Walt Disney Studios motion picture was second in list in number of movies it was top in the list of highest collector distributor. Walt Disney Studios motion picture has collected over 82 billion dollars and Warner bros being on the second position had earned over 63 billion dollars. The stats clearly indicates that the money made by Walt Disney Studios motion picture was way higher than other Distributors.

If we are to talk about the earning of every year that has come in the list of movies, the following line graph shows how much they have made throughout the year.



In the year 2019 they had the most earning comparing to other years and the second and third was 2018 and 2017 respectively. For the further analysis I found that the movies which were in ['Action', 'Adventure', 'Sci-Fi'] genre and it has earned more in compare to others.

The movies which were in the PG-13 category was tend to be earning well in compare to others. It is also seeming to have the data that points that Walt Studios Motion Pictures and Warner Bros have been producing the movie by following the same kind of pattern that made them successful and better than the rest.

Another analysis states that the movies which earn well in Domestic seems to have greater effect on global sales as well.

What's Next?

Nowadays, Since the pandemic has happened, there has been certain raise in the number of platforms where we can have newly released movies and web series as well. Some of the movies gets directly released on OTT platform which is now in play. So, the distributors have made the possibility for higher sales since its online people will watch it on multiple occasion and which will increase sales. So, I hope to get a dataset on sales of Ott platform hits and do analysis on that. I have made a dashboard also in Tableau which will be posted in the folder above to have a better understanding of insights.

